Generating Language Activities in Real-Time for English Learners using Language Muse

Jill Burstein
Educational Testing Service
Princeton, USA
jburstein@ets.org

Nitin Madnani
Educational Testing Service
Princeton, USA
nmadnani@ets.org

John Sabatini
Educational Testing Service
Princeton, USA
jsabatini@ets.org

Dan McCaffrey
Educational Testing Service
Princeton, USA
dmccaffrey@ets.org

Kietha Biggers
Educational Testing Service
Princeton, USA
kbiggers@ets.org

Kelsey Dreier
Educational Testing Service
Princeton, USA
kdreier@ets.org

ABSTRACT
K-12 education standards in the U.S. require all students to read complex texts across many subject areas. The *Language Muse*™ *Activity Palette* is a web-based language-instruction application that uses NLP algorithms and lexical resources to automatically generate language activities and support English language learners' content comprehension and language skills development. The system's online platform for activity generation, scoring, and feedback is scalable for MOOCs, as well as for other online learning settings.

INTRODUCTION
The Common Core State Standards adopted by most U.S states explicitly emphasize the need for students to read complex subject area texts to prepare for college and careers [6]. Classroom texts may contain language unfamiliar to English language learners (ELLs), e.g. figurative language. ELLs could be disadvantaged without scaffolding to aid in comprehension of unfamiliar language [8]. One way to help is through the use of linguistic activities designed to get ELLs familiar with the language used in the subject area texts [7].

We present the Language Muse Activity Palette (*LM* heretofore), an open-access, web-based tool [1], and discuss a small-scale instructional pilot intervention that has shown promise with regard to addressing ELL content comprehension and language learning needs. Teachers can input their own classroom texts into LM to automatically generate language activities in real-time which can then be assigned to students online. The activities are generated using several existing NLP algorithms and lexical resources designed to help ELLs with multiple aspects of language learning needed to support content comprehension: vocabulary, syntactic structures, and discourse structure.

LM is related to existing NLP work on automatic question generation [2, 12, 14]. In contrast to previous work, it can generate over 20 activity types for any given classroom text, covering a large set of language constructs, and offers activity customizability. In addition, many activities can be automatically scored. Analytics can also be generated for students' language proficiency from both automatically-scored and teacher-scored activities.

Below, we describe the LM NLP backend, the teacher and student interactions, findings from a small-scale instructional intervention, and future work.

NLP BACKEND
LM relies on a backend that uses NLP algorithms to identify linguistic features contained in an input text [5]. These features include: (a) lexical entities (single word and multi-word expressions), (b) syntactic structures, and (c) discourse relations. LM also relies on a few manually-crafted resources either directly, or indirectly as a filter for statistical NLP algorithms that may yield somewhat noisier outputs. This limits the teachers' need to edit over-generated, incorrect options.

Lexical resources are used for activities related to these language elements: homonyms [3], cognates [5], academic words [9], and antonyms [10]. Synonym-based activities are powered using a thresholded combination of WordNet, a distributional thesaurus [13], and statistically extracted paraphrases [1]. Multiword expression activities are generated using a rank-ratio based collocation detection algorithm trained on the Google Web1T *n*-gram corpus [11]. Regular expressions defined on constituency parses generate phrasal and sentential structures for activities related to contractions, complex verb, noun phrases, relative clauses, and multi-clause sentences. A morphological analyzer and word-form database are used to generate activities related to derivational and inflectional word forms. Discourse relations related to cause-effect, compare-

[1]available at http://languagemuse.10clouds.com.

L@S 2017, April 20 - 21, 2017, Cambridge, MA, USA

ISBN 978-1-4503-4450-0/17/04. . . $15.00
DOI: http://dx.doi.org/10.1145/3051457.3053988

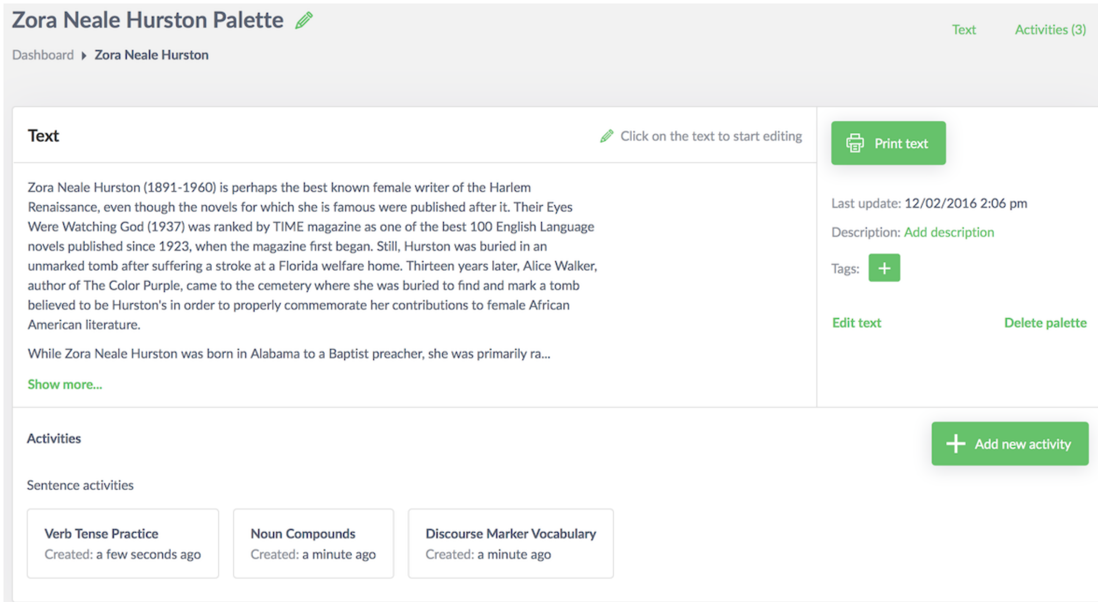

Figure 1. An example activity palette generated using Language Muse.

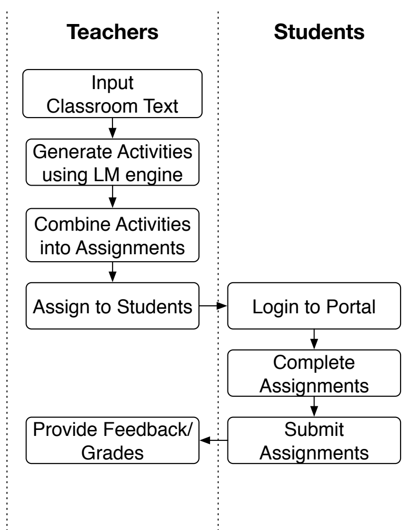

Figure 2. A flowchart illustrating the complete Language Muse instructional workflow from activity generation to students receiving feedback on their work.

contrast, and evidence draw from an adapted rule-based, discourse analyzer [4].

PALETTE & ASSIGNMENT CREATION, AND SCORING

Teachers upload a classroom text into LM. The engine automatically generates over 20 activities based on linguistic features identified in the text. Teachers then select activities to create an "activity palette" (Figure 1) — a set of text-specific activities — to support one or more learning objectives, such as "practice with derivational word forms".

Full palettes, or specific activities in a palette, can be used to create assignments targeting the learning objective. Assignments can be administered to and completed online by students. Multiple choice, and cloze activities are automatically scored. For activities requiring open-ended responses, teachers provide scores and written feedback. Teachers and students may view scores and feedback at any time. See Figure 2 for a high-level overview of this workflow.

INSTRUCTIONAL INTERVENTION

In Spring 2016, a 6-week instructional intervention study was conducted to examine the promise and feasibility of LM use in a classroom setting, in preparation for a large-scale randomized control trial (RCT) in Spring 2017.

- **Teacher & Student Participants**. Results are based on 12 English Language Arts (ELA), Science, and Social Studies teachers from two participating middle schools where ELL populations were over 33%. 167 students who completed pre- and post-tests were included in the analysis presented here.

- **Instruments**. The RISE reading assessment [15] was administered to students pre- and post-intervention to examine intervention outcomes. The test contains 6 component measures related to reading proficiency, e.g., vocabulary and morphology. An observational protocol [5] was used to collect teacher observation data pre-, during and post-intervention. Teacher perception surveys were administered post-intervention.

- **Preliminary Findings**. The pre-post assessment outcomes were difficult to interpret, especially outside of an RCT. Gains were observed in some components, but there were also score losses – potentially due to motivation, since the assessments were no-stakes. Observational data suggested

Activity Type	Count
Academic Vocabulary	19
Antonyms	9
Cognates (Spanish)	17
Compare/Contrast	6
Finding Homonyms	17
Multiple Clauses	16
Phrasal Verbs	3
Referential Terms	5
Summary Practice	3
Synonyms	18
Variant Word Forms	21
Verb Tenses	12
Word Stems	9

Table 1. Teachers' self-reported usage showing counts of activities during the instructional intervention.

that teachers productively integrated LM into classroom instruction. A positive survey finding showed that all teachers successfully completed the intervention, and reported that activities completely (47%) or mostly (39%) fulfilled intended learning objectives. See Table 1 for activity usage during the intervention.

FUTURE WORK

In Spring 2017, an instructional intervention will be conducted with LM in an RCT with approx. 20 U.S. middle schools with high EL populations. Outcomes showing promising use of and positive reactions to LM, and pre-post assessment gains would suggest promise of LM as a classroom tool, with potential applications in language learning MOOCs and other online classrooms to support learning at scale. Currently, Paper Airplanes[2] is also exploring LM use for one-on-one online English tutoring for Syrian students.

ACKNOWLEDGMENTS

Research presented in this paper was supported by the Institute of Education Science, U.S. Department of Education, Award Number R305A140472, and by the 10Clouds front-end development team.

REFERENCES

1. Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with Bilingual Parallel Corpora. In *Proceedings of ACL*.

2. Jonathan C Brown, Gwen A Frishkoff, and Maxine Eskenazi. 2005. Automatic Question Generation for Vocabulary Assessment. In *Proceedings of EMNLP*.

3. Jill Burstein, Martin Chodorow, and Claudia Leacock. 2004. Automated Essay Evaluation: The Criterion Online Writing Service. *AI Magazine* 25, 3 (2004), 27.

4. Jill Burstein, Karen Kukich, Susanne Wolff, Chi Lu, and Martin Chodorow. 1998. Enriching Automated Scoring using Discourse Marking. In *Proceedings of the ACL Workshop on Discourse Relations and Discourse Marking*.

5. J. Burstein, J. Shore, J. Sabatini, B. Moulder, J. Lentini, K. Biggers, and S. Holtzman. 2014. From Teacher Professional Development to the Classroom. *Journal of Educational Computing Research* 51, 1 (2014), 119–144.

6. CCSSO. 2010. *Common Core State Standards for English language Arts & Literacy in History/Social Studies, Science, and Technical Subjects. Appendix A: Research supporting key elements of the Standards*. Washington, DC.

7. Frances Christie. 1989. *Language Education*. Oxford University Press, Oxford, UK.

8. Rhonda Coleman and Claude Goldenberg. 2012. The Common Core Challenge for English Language Learners. *Principal Leadership* (2012), 46–51.

9. A. Coxhead. 2000. The Academic Word List. Retrieved from `http://www.victoria.ac.nz/lals/resources/academicwordlist/`. (2000).

10. Christiane Fellbaum. 1998. *WordNet*. Blackwell Publishing Ltd.

11. Yoko Futagi, Paul Deane, Martin Chodorow, and Joel Tetreault. 2008. A Computational Approach to Detecting Collocation Errors in the Writing of Non-native Speakers of English. *Computer Assisted Language Learning* 21, 4 (2008).

12. Michael Heilman and Noah A Smith. 2010. Good question! Statistical Ranking for Question Generation. In *Proceedings of NAACL*.

13. Dekang Lin. 1998. Automatic Retrieval and Clustering of Similar Words. In *Proceedings of COLING*.

14. Ruslan Mitkov and Le An Ha. 2003. Computer-aided Generation of Multiple-choice Tests. In *Proceedings of the Workshop on Building Educational Applications*.

15. John Sabatini, Kelly Bruce, Jonathan Steinberg, and Jonathan Weeks. 2015. *SARA Reading Components Tests, RISE Forms: Technical Adequacy and Test Design, 2nd Edition*. Technical Report ETS-RR-15-32.

[2] `http://www.paper-airplanes.org/`